Unlock the power of your organization’s hidden data

Use advanced natural language processing and content mining to transform business information

Decisions grounded in data drive modern business. Organizations rely on ongoing analysis of the information they possess to make rational choices about purchasing, hiring, inventory control, resource allocation and more. To make the best-informed decisions, merely possessing data isn’t enough—an organization can use only the information it can meaningfully access and analyze. That means more data is not necessarily more useful. Businesses can be overwhelmed with information that is only potentially useful.

Ninety percent of the world’s data has been created in the last two years,¹ and the trend is only accelerating. Deepening the challenge, an estimated 80 percent of enterprise data—sometimes more—is locked into natural language notes stored as plain text, scanned documents, video or other forms that keep it from ready analysis.² This valuable but untapped data spans emails, web-form data, repair and failure reports, patents, social media posts and more. These non-parallel data types can keep valuable data trapped beneath the surface. As data sources and data volumes continue to rise, data science has become a vital tool in wrangling disparate data sources, turning abstract data into actionable predictive information and, ultimately, into business decisions.

Highlights

- Transform unstructured data into powerful, searchable structured data
- Analyze unstructured data to discover key trends and actionable insights
- Empower informed business decision making by consolidating data sources into a unified search pool
- Enable robust searching and seamless information sharing with the intuitive interface of IBM® Data Science Experience
- Simplify access to industry-standard data science analysis tools
IBM Data Science Experience can be extended with the natural language processing and content mining capabilities of IBM Watson® Explorer Deep Analytics Edition to help unlock the value of the unstructured data that businesses possess, but cannot effectively mine for insights.

Why unstructured data matters to business

Only a fraction of the data that companies have at their disposal for making business decisions is structured—that is, held in a clean, regular form that conforms to the formats databases require for it to be readily searchable. A database containing only telephone numbers and associated names, all identically formatted, illustrates structured data.

Unstructured and semi-structured data includes everything else. Crucially for businesses, this includes much of the text-based information under their control—but it lacks the regularity and predictability of structured data. From plain-text notes or Voice of Customer data to medical reports, unstructured or semi-structured data makes up some of the most valuable information a business may hold, if the trends and insights associated with it can be surfaced and understood.

Finding insights in expanding data stores

All varieties of unstructured and semi-structured data contrast with the contents of well-ordered, carefully constructed databases. Sales data, expenditures and more may be kept in just such data stores. But while a database can be readily sorted for the answers relevant to your organization, it is highly time-consuming to explore, discover patterns, and gain actionable insights from unstructured data.

Furthermore, for large volumes of unstructured data, human expertise simply cannot scale to understand the underlying trends and insights. This means that help from advanced analytics is vital if businesses are to make sense of all their data to create value and grow.

Simplifying access to powerful tools

Data science has conventionally been accomplished with command-line tools. Many of these are highly domain-specific programming languages, such as Python and R, that have been designed or adapted for efficiently sorting and sifting large volumes of data. Also key to data science is the template-driven textual analysis made possible by IBM SPSS®. Both kinds of software are powerful, but in isolation, they are limited in their application because their use requires specialized expertise and carefully tailored data. Enterprise tools that expand the pool of data to include unstructured data sources, enabling natural language processing for intuitive searches, give a broader group of users the power of detailed search and insight gathering.
Data Science Experience is an advanced, extensible environment that provides data scientists with a unified, user-friendly portfolio of analytics tools. The IBM solution offers a visual approach suitable for business-oriented data scientists, along with collaboration interfaces for sharing data among peer users. Because specialized add-ons can extend Data Science Experience, its capabilities can be tuned to the needs of each organization.

A powerful example: Watson Explorer provides powerful natural language processing with content mining capabilities, so unstructured and semi-structured textual data can be can be analyzed by the Data Science Experience platform.

**Business benefits of full data visibility**

Data analytics means nothing to a business without context. All data analysis should help an organization understand its current state, predict the future, and change outcomes, with the least possible expenditure of valuable personnel time. Watson Explorer uses natural language processing and content mining to pull data together, analyze data for trends and patterns, and extract actionable insights. The process helps to enable confident business decisions while minimizing busywork.

When unstructured or semi-structured data has been transformed into actionable information with Watson Explorer, that information can be examined in the same context as other structured data sources. Once standardized, all of an organization’s data can be explored with both visual and command-line data-analysis tools available through Data Science Experience. This “single pane of glass” approach to understanding data in context can eliminate both guesswork and effort.

Organizations currently employing at the individual or department level IBM SPSS Text Analytics, available separately from IBM, can use the power of Watson Explorer and Data Science Experience to help amplify the power of their data to enterprise levels. SPSS Text Analytics uses robust domain- or industry-specific templates for interpreting text-based data. Facts extracted from previously unstructured data using Data Science Experience can be entered into IBM SPSS Text Analytics to help improve an organization’s native statistical models. Alternatively, Data Science Experience can be configured to push the transformed data directly into an IBM SPSS Modeler node to improve existing SPSS models.
The exploration and insight generation made possible with Data Science Experience means that previously “dark” data can be used to spot sales trends, identify complaints or failures, or help evaluate business processes. That means you can base business moves—for example, to increase customer satisfaction and enhance your competitive advantage—with cognitive insights drawn from all your data, not just the 20 percent in readily searchable database formats.

Data Science Experience users can use Jupyter notebooks to consolidate and share data—and the code used to process the data, visualizations of the results, and text—as well as rich media explanations of the notebook’s numerical content. With the integration of Watson Explorer with Data Science Experience, the natural language processing and content mining capabilities that Watson Explorer brings can be invoked directly from these notebooks through an application programming interface (API).

**Implementing the power of Watson Explorer for Data Science Experience**

To add the power of Watson Explorer to an existing Data Science Experience environment requires only the installation and configuration of the Watson Explorer add-on. Configuration includes the enablement of Data Science Experience to manage data collection and other capabilities of Watson Explorer.

With its intuitive user interface, Watson Explorer simplifies content mining with natural language processing for context-appropriate results. Data scientists can use the content mining to quickly build custom cognitive solutions.

To enable its natural language query capabilities, stores of unstructured data must then be defined and linked to the solution, so their contents can be incorporated. For useful output, data scientists define the data sought and supply appropriate search parameters.
Putting the power of all your data to work

The insights available after transforming scattered source content into structured data apply across a broad range of organizational decisions.

- In the field of insurance, disparate data sources are a given. Even in the case of simple automobile accident data, formerly unstructured data such as narrative description of the collision, scanned police reports, telephone transcripts and correspondence with multiple insurance companies can be consolidated. It then can be accessed—along with initially structured numeric information such as car-sensor data—to construct a complete story of the event.

- For manufacturing operations, unstructured data can contain valuable insights into quality and efficiency. Hidden in daily reports or operation logs, customer emails, and feedback forms is information that may reflect defect rates, geographic trends and production delays by lot number or location.

- Consumer product manufacturers rely on consumer reputation. The sentiment analysis that Watson Explorer makes possible can be readily accessed through the Data Science Experience platform to understand trends in reviews or complaints. These insights are fully possible only when all forms of textual data are included.

- Government applications for insight gained from initially unstructured data include efforts to increase traffic safety by analyzing accident reports, allocating policing efforts to reflect public needs, and spotting possible financial mismanagement or budget misallocations.

A Japanese health insurance company used Watson Explorer as part of an effort aimed at improving health insurance payment decisions by identifying important but ambiguous medical information in medical certificates. The result: 20 percent fewer mistakenly unpaid claims, with greater than 90 percent accuracy in identifying and coding medical terms—and a reduction in assessment workload of more than 30 percent.3

Why IBM?

Watson Explorer embodies decades of IBM research and expertise in offering powerful natural language processing models that can be configured, trained, and evaluated easily, enabling advanced content analytics and search. Organizations across industries are using the natural language processing and content mining capabilities of IBM Watson Explorer to make important business decisions based on the actionable insights gained. Watson Explorer has demonstrated successful outcomes across industries including banking and finance, insurance, manufacturing and government. This cross-domain experience, backed by the efforts of IBM Research, helps ensure leading-edge product development and support for users.
For more information
To learn more about how the integration of IBM Data Science Experience with IBM Watson Explorer can help your organization capture and leverage its holdings of unstructured data to drive effective business decisions, please contact your IBM representative or IBM Business Partner, or visit the following website: ibm.biz/ContentAnalytics

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: ibm.com/financing

© Copyright IBM Corporation 2018
IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
July 2018

IBM, the IBM logo, ibm.com, Watson, and SPSS are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NONINFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

© Copyright IBM Corporation 2018
IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
July 2018

IBM, the IBM logo, ibm.com, Watson, and SPSS are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NONINFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

1 “Big Data, for better or worse: 90% of world’s data generated over last two years,” Science Daily, May 22, 2013. www.sciencedaily.com/releases/2013/05/130522085217.htm


3 Based on IBM customer experience.